CHAPTER 13

Gateways

Revised: September 30, 2011; OL-21733-18

Gateways provide a number of methods for connecting an IP telephony network to the Public Switched Telephone Network (PSTN), a legacy PBX, or key systems. Gateways range from specialized, entry-level and stand-alone voice gateways to high-end, feature-rich integrated router and Cisco Catalyst gateways.

This chapter explains important factors to consider when selecting a Cisco gateway to provide the appropriate protocol and feature support for your IP Telephony network. The main topics discussed in this chapter include:

What's New in This Chapter

Table 13-1 lists the topics that are new in this chapter or that have changed significantly from previous releases of this document.

Table 13-1 New or Changed Information Since the Previous Release of This Document

New or Revised Topic	Described in	Revision Date
Gateway configuration examples were removed from this document but can be found in the *Cisco IOS Fax, Modem, and Text Support over IP Configuration Guide*, which is available at http://www.cisco.com/en/US/docs/ios/voice/fax/configuration/guide/12_4t/vf_12_4t_book.html	Gateway Configuration Examples, page 13-26	September 30, 2011
Video telephony gateways	Gateways for Video Telephony, page 13-27	November 15, 2010
Fax and modem support	Fax and Modem Support, page 13-19	April 2, 2010

Traffic Patterns and Gateway Sizing

This section presents a high-level discussion of the differences between various traffic models or patterns and how they can affect voice gateway selection. The emphasis is on traffic patterns and gateway sizing for traffic-intensive deployments.

Definitions and Terminology

This section uses the following terms and definitions:

- Simultaneous calls

  The number of calls that are all active in the system at the same time.

- Maximum simultaneous calls

  The maximum number of simultaneous calls in active (talk) state that the system can handle. The number of calls expected to be active simultaneously during the *busy hour* of the day should not exceed this number.

- Calls per second (cps)

  The call arrival rate, described as the number of calls that arrive (that is, new call setup attempts) in one second. Call arrival rates are also often quoted in calls per hour, but this metric is looser in the sense that 100 calls arriving in the last five seconds of an hour provides an average call arrival rate of 100 calls per hour (which is an extremely low rate for a communications system), while it also provides an arrival rate of 20 calls per second (which is a high rate). Sustaining 20 calls per second for an entire hour would result in 72,000 calls per hour. Therefore calls-per-hour is not a very useful metric for ascertaining a system's ability to handle bursty call arrival traffic patterns.

- Busy Hour Call Attempts (BHCA)

  The number of calls attempted during the busiest hour of the day (the peak hour). This is the same as the calls-per-second rating for the busiest hour of the day, but it is expressed over a period of an hour rather than a second. For example, 10 cps would be equal to 36,000 calls per hour. There is also a metric for Busy Hour Call Completions (BHCC), which can be lower than the BHCA (call attempts) under the assumption that not all calls are successful (as when a blocking factor exists). This chapter assumes 100% call completions, so that BHCA = BHCC.

- Bursty traffic

  Steady arrival means the call attempts are spaced more or less equally over a period of time. For example, 60 calls per hour at a steady arrival rate would present one call attempt roughly every minute (or approximately 0.02 cps). With bursty arrival, the calls arriving over a given period of time (such as an hour) are not spaced equally but are clumped together in one or more spikes. In the worst case, an arrival rate of 60 calls per hour could offer all 60 calls in a single second of the hour, thus averaging 0 cps for most of the hour with a peak of 60 cps for that one second. This kind of traffic is extremely stressful to communications systems.

- Hold time

  This is the period of "talk time" on a voice call; that is, the period of time between call setup and call teardown when there is an open speech path between the two parties. A hold time of 3 minutes (180 seconds) is an industry average used for traffic engineering of voice systems. The shorter the hold time on the average call, the greater the percentage of system CPU time spent on setting up and tearing down calls compared to the CPU time spent on maintaining the speech path.

PSTN Traffic Patterns

Traffic, when used in the context of voice communication systems, refers to the volume of calls being sent and/or received. Of particular importance is the traffic carried by external circuits such as the public switched telephone network (PSTN). Traffic is measured in Erlangs, and an Erlang is defined as one call lasting for one hour. This section does not go into any further detail on Erlangs other than to say that there are mathematical tables (Erlang-B and Erlang-C) that are used to calculate how many circuits are required for a given amount of offered traffic.

The amount of traffic received and generated by your business determines the size of the external circuits required. However; many customers typically continue to use the same number of circuits for their IP-based communications system as they previously used for a TDM-based system. While this sizing method might work if no issues are encountered, the process of ongoing system traffic analysis should be part of any routine maintenance practices. Traffic analysis can show that the system is over-provisioned for the current levels of traffic (and, therefore, the customer is paying for circuits that are not needed) or, conversely, that the system is under-provisioned and may be suffering from occasional blocked and/or lost calls, in which case increasing the number of circuits will remedy the situation.

Normal Business Traffic Profile

Most customers have a normal traffic profile, which means that they typically have two *busy hours* per day, one occurring during the morning from 10:00 to 11:00 and the other in the afternoon from 14:00 to 15:00. These busy-hour patterns can often be attributed to such things as employees starting the work day or returning from a lunch break. The calls themselves tend to have longer hold times and they tend to arrive and leave in a steady manner. A generally accepted industry average holding time to use for traffic calculations is 3 minutes.

Assuming that the communications system is engineered with the busy-hour traffic in mind, no issues should arise. Engineering a system below these levels will result in blocked and/or lost calls, which can have a detrimental effect on business.

Contact Center Traffic Profile

Contact centers present somewhat different patterns of traffic in that these systems typically handle large volumes of calls for the given number of agents or interactive voice response (IVR) systems available to service them. Contact centers want to get the most out of their resources, therefore their agents, trunks, and IVR systems are kept busy all the while they are in operation, which usually is 24 hours a day. Call queuing is typical (when incoming call traffic exceeds agent capacity, calls wait in queue for the next available agent), and the agents are usually dedicated during their work shifts to taking contact center calls.

Call holding times in contact centers are often of a shorter average duration than normal business calls. Contributing to the shorter average call holding time is the fact that many calls interact only with the IVR system and never need to speak to a human agent (also termed self-service calls). A representative holding time for self-service calls is about 30 seconds, while a call that talks to an agent has an average holding time of 3 minutes (the same as normal business traffic), making the overall average holding time in the contact center shorter than for normal business traffic.

The goal of contact centers to optimize resource use (including IVR ports, PSTN trunks, and human agents), combined with the fact that contact centers are systems dedicated to taking telephone calls, also presents the system with higher call arrival rates than in a typical business environment. These call arrival rates can also peak at different times of day and for different reasons (not the usual busy hour) than normal business traffic. For example, when a television advertisement runs for a particular holiday

package with a 1-800 number, the call arrival rate for the system where those calls are received will experience a peak of traffic for about 15 minutes after the ad airs. This call arrival rate can exceed the average call arrival rate of the contact center by an order of magnitude.

Gateway Sizing for Contact Center Traffic

Short call durations as well as bursty call arrival rates impact the PSTN gateway's ability to process the traffic. Under these circumstances the gateway needs more resources to process all calls in a timely manner, as compared to gateways that receive calls of longer duration that are presented more uniformly over time. Because gateways have varying capabilities to deal with these traffic patterns, careful consideration should be given to selecting the appropriate gateway for the environment in which it will operate. Some gateways support more T1/E1 ports than others, and some are more able than others to deal with multiple calls arriving at the same time.

For a traffic pattern with multiple calls arriving in close proximity to each other (that is, high or bursty call arrival rates), a gateway with a suitable rating of calls per second (cps) is the best fit. Under these conditions, using calls with 15-second hold times, the Cisco AS5400XM Universal Gateway can maintain 16 cps with 250 calls active at once, the Cisco 3845 Integrated Services Router can maintain 13 cps with 200 calls active at once, and the Cisco 3945 Integrated Services Router can maintain 28 cps with 420 calls active at once. The performance of the Cisco AS5350XM Universal Gateway is identical to that of the AS5400XM in terms of calls per second.

For traffic patterns with a steady arrival rate, the maximum number of active calls that a gateway can handle is generally the more important consideration. Under these conditions, using calls with 180-second hold times, the Cisco AS5400XM Universal Gateway can maintain 600 simultaneously active calls with a call arrival rate of up to 3.3 cps, the Cisco 3845 Integrated Services Router can maintain 450 simultaneously active calls with a call arrival rate of up to 2.5 cps), and the Cisco 3945 Integrated Services Router can maintain 720 simultaneously active calls with a call arrival rate of up to 4 cps).

These numbers assume that all of the following conditions apply:

- CPU utilization does not exceed 75%.
- PSTN gateway calls are made with ISDN PRI trunks using H.323.
- Real Time Control Protocol (RTCP) timer is set to the default value of 5 seconds.
- Voice Activity Detection (VAD) is off.
- G.711 uses 20 ms packetization.
- Cisco IOS Release 15.0.1M is used.
- Dedicated voice gateway configurations are used, with ethernet (GE) egress and no QoS features. (Using QoS-enabled egress interfaces or non-ethernet egress interfaces, or both, will consume additional CPU resources.)
- No supplementary call features or services are enabled – such as general security (for example, access control lists or firewalls), voice-specific security (TLS, IPSec and/or SRTP), AAA lookups, gatekeeper-assisted call setups, VoiceXML or TCL-enabled call flows, call admission control (RSVP), and SNMP polling/logging. Such extra call features will use additional CPU resources.

Voice Activity Detection (VAD)

VAD is a digital signal processing feature that suppresses the creation of most of the IP packets during times when the speech path in a particular direction of the call is perceived as being silent. Typically only one party on a call speaks at a time, so that packets need flow in only one direction, and packets in

the reverse (or silent) direction need not be sent except as an occasional keepalive measure. VAD can therefore provide significant savings in the number of IP packets sent for a VoIP call, and thereby save considerable CPU cycles on the gateway platform. While the actual packet savings that VAD can provide varies with the call flow, the application, and the nature of speaker interactions, it tends to use 10% to 30% fewer packets than would be sent for a call made using a VAD-off configuration.

VAD is most often turned off in endpoints and voice gateways deployed in Unified CM networks; VAD is most often turned on in voice gateways in other types of network deployments.

Codec

Both G.711 and G.729A use as their default configuration a 20 ms sampling time, which results in a 50 packets per second (pps) VoIP call in each direction. While a G.711 IP packet (200 bytes) is larger than a G.729A packet (60 bytes), this difference has not proven to have any significant effect on voice gateway CPU performance. Both G.711 and G.729 packets qualify as "small" IP packets to the router, therefore the packet rate is the salient codec parameter affecting CPU performance.

Performance Overload

Cisco IOS is designed to have some amount of CPU left over during peak processing, to handle interrupt-level events. The performance figures in this section are designed with the processor running at an average load of approximately 75%. If the load on a given Cisco IOS gateway continually exceeds this threshold, the following will result:

- The deployment will not be supported by Cisco Technical Assistance Center (TAC).
- The Cisco IOS Gateway will display anomalous behavior, including Q.921 timeouts, longer post-dial delay, and potentially interface flaps.

Cisco IOS Gateways are designed to handle a short burst of calls, but continual overloading of the recommended call rate (calls per second) is not supported.

Note With any gateway, you might be tempted to assign unused hardware ports to other tasks, such as on a CMM gateway where traffic calculations have dictated that only a portion of the ports can be used for PSTN traffic. However, the remaining ports *must* remain unused, otherwise the CPU will be driven beyond supported levels.

Performance Tuning

The CPU utilization of a Cisco IOS Voice gateway is affected by every process that is enabled in a chassis. Some of the lowest level processes such as IP routing and memory defragmentation will occur even when there is no live traffic on the chassis.

Lowering the CPU utilization can help to increase the performance of a Cisco IOS Voice Gateway by ensuring that there are enough available CPU resources to process the real-time voice packets and the call setup instructions. Some of the techniques for decreasing CPU utilization are described in Table 13-2.

Table 13-2 Techniques for Reducing CPU Utilization

Technique	CPU Savings	Description
Enable Voice Activity Detection (VAD)	Up to 20%	Enabling VAD can result in up to 45% fewer voice packets in typical conversations. The difficultly is that, in scenarios where voice recognition is used or there are long delays, a reduction in voice quality can occur. Voice appears to "pop" in at the beginning and "pop" out at the end of talk spurts.
Disable Real Time Control Protocol (RTCP)	Up to 5%	Disabling RTCP results in less out-of-band information being sent between the originating and terminating gateways. This results in lower quality of statistics displayed on the paired gateway. This can also result in the terminating gateway having a call "hang" for a longer period of time if RTCP packets are being used to determine if a call is no longer active.
Disable other non-essential functions such as: Authentication, Authorization, and Accounting (AAA); Simple Network Management Protocol (SNMP); and logging	Up to 2%	Any of these processes, when not required, can be disabled and will result in lower CPU utilization by freeing up the CPU to provide faster processing of real-time traffic.
Change call pattern to increase the length of the call (and reduce the number of calls per second)	Varies	This can be done by a variety of techniques such as including a long(er) introduction prompt played at the beginning of a call or adjusting the call script at the call center.

Additional Information

For more information on Cisco Voice Gateway capabilities and call center traffic analysis, refer to the following sources:

- Cisco Voice Gateway Solutions:

  http://www.cisco.com/en/US/products/sw/voicesw/index.html#~all-prod

- Gateway protocols supported with Cisco Unified Communications Manager (Unified CM):

  http://www.cisco.com/en/US/docs/voice_ip_comm/cucm/admin/8_0_1/ccmsys/a08gw.html

- Interfaces and signaling types supported by the following Cisco Voice Gateways:
  - Cisco 3900 Series Integrated Services Routers

    http://www.cisco.com/en/US/products/ps10536/products_relevant_interfaces_and_modules.html
  - Cisco 2900 Series Integrated Services Routers

    http://www.cisco.com/en/US/products/ps10537/products_relevant_interfaces_and_modules.html
  - Cisco 3800 Series Integrated Services Routers

    http://www.cisco.com/en/US/products/ps5855/products_relevant_interfaces_and_modules.html
  - Cisco 2800 Series Integrated Services Routers

    http://www.cisco.com/en/US/products/ps5854/products_relevant_interfaces_and_modules.html

- Gateway features supported with MGCP, SIP, and H.323:

  http://www.cisco.com/en/US/prod/collateral/routers/ps259/product_data_sheet0900aecd8057f2e0.pdf

- SIP gateway RFC compliance:

  http://www.cisco.com/en/US/prod/collateral/voicesw/ps6790/gatecont/ps6831/product_data_sheet0900aecd804110a2.html

- Skinny Client Control Protocol (SCCP) feature support with FXS gateways:

  http://www.cisco.com/en/US/prod/collateral/voicesw/ps6790/gatecont/ps2250/ps5516/product_data_sheet09186a00801d87f6.html

- Gateway capacities and minimum releases of Cisco IOS and Unified CM required for conferencing, transcoding, media termination point (MTP), MGCP, SIP, and H.323 gateway features:

  http://www.cisco.com/en/US/prod/collateral/routers/ps259/product_data_sheet0900aecd8057f2e0.pdf

- Various voice traffic calculators, including Erlang calculators:

  http://www.erlang.com/calculator/

Considerations for Gateway Redundancy

When deploying a gateway solution, give careful consideration to redundancy when compared with scalability. For example, the Cisco VGD 1T3 Voice Gateway is capable of delivering 28 T1 lines in one physical circuit. However; considering the inherent need for redundancy with PSTN services, multiple smaller gateways delivering the same overall physical quantity of service may be more appropriate. Multiple gateways further allows for placement in different physical locations, thus allowing for another level of redundancy, in this case spatial redundancy.

Gateway deployments involving contact with emergency services must also be considered, and sometimes more than one solution may be necessary. For example, consider a small branch location connected to the PSTN through a SIP trunk located at a remote headquarters. If there is either a WAN or SIP trunk failure, the branch location must still be able to contact the emergency services. In this case the best solution would be either a local analog or PRI service (that is, either a standalone analog service or a PRI service terminated on the branch router).

TDM and VoIP Trunking Gateways

Until approximately 2006, the only choice for an enterprise to connect its internal VoIP network to voice services outside the enterprise was via TDM gateways to the traditional PSTN. Cisco offers a full range of TDM gateways with analog and digital connections to the PSTN as well as to PBXs and key systems. TDM connectivity covers a wide variety of low-density analog (FXS and FXO), low density digital (BRI), and high-density digital (T1, E1, and T3) interface choices.

Starting around 2006, new voice service options to an enterprise started to become available from service providers, most often referred to as SIP trunk services. Using a SIP trunk for connecting to PSTN and other destinations outside the enterprise involves an IP-to-IP connection at the edge of the enterprise's VoIP network. The same functions traditionally fulfilled by a TDM gateway are still needed at this interconnect point, including demarcation, call admission control, ensuring QoS, a troubleshooting boundary, security checks, and so forth. For SIP trunking connections, the Cisco Unified Border Element fulfils these functions as a session border controller (SBC) at the interconnect point between the

enterprise and the service provider network. Cisco Unified Border Element also performs protocol translation functions to interconnect H.323 and SIP equipment, or to interconnect SIP equipment using different variations of SIP implementations. Cisco Unified Border Element can also perform transcoding. If used for one of these functions, Cisco Unified Border Element may also be used internal to the enterprise network at interconnect points between equipment that cannot interoperate without a protocol translation or transcoding service.

TDM gateway platforms are discussed in detail in the remainder of this chapter. Cisco Unified Border Element is discussed in greater detail in the chapter on Cisco Unified CM Trunks, page 14-1. Both functions can be enabled on the same Cisco Integrated Services Router (ISR) platform at the same time.

Understanding Cisco Gateways

Cisco access gateways enable Cisco Unified Communications Manager (Unified CM) to communicate with non-IP telecommunications devices. There are two types of Cisco access gateways, analog and digital.

Cisco Access Analog Gateways

There are two categories of Cisco access analog gateways, trunk gateways and station gateways.

- Access analog station gateways

  Analog station gateways connect Unified CM to Plain Old Telephone Service (POTS) analog telephones, interactive voice response (IVR) systems, fax machines, and voice mail systems. Station gateways provide Foreign Exchange Station (FXS) ports.

- Access analog trunk gateways

  Analog trunk gateways connect Unified CM to PSTN central office (CO) or PBX trunks. Trunk gateways provide Foreign Exchange Office (FXO) ports for access to the PSTN, PBXs, or key systems, and E&M (recEive and transMit, or ear and mouth) ports for analog trunk connection to a legacy PBX. Whenever possible, use digital gateways to minimize any answer and disconnect supervision issues. Analog Direct Inward Dialing (DID) and Centralized Automatic Message Accounting (CAMA) are also available for PSTN connectivity.

Cisco Access Digital Trunk Gateways

A Cisco access digital trunk gateway connects Unified CM to the PSTN or to a PBX via digital trunks such as Primary Rate Interface (PRI), Basic Rate Interface (BRI), or T1 Channel Associated Signaling (CAS). Digital T1 PRI trunks may also be used to connect to certain legacy voice mail systems.

Tuning Gateway Gain Settings

Connecting a Cisco Unified Communications network to the PSTN through gateways requires that you properly address voice quality issues arising from echo and signal degradation due to power loss, impedance mismatches, delay, and so forth. For this purpose, you must establish a Network Transmission Loss Plan (NTLP), which provides a complete picture of signal loss in all expected voice paths. Using this plan, you can identify locations where signal strength must be adjusted for optimum loudness and effective echo cancellation. Note that not all carriers use the same loss plan, and that the

presence of cellular networks adds further complexity in creating the NTLP. Cisco does not recommend adjusting input gain and output attenuation on gateways without first completing such an NTLP. For more information, refer to *Echo Analysis for Voice Over IP*, available at

http://www.cisco.com/en/US/docs/ios/solutions_docs/voip_solutions/EA_ISD.pdf

Gateway Selection

When selecting an IP telephony gateway, consider the following factors:

- Core Feature Requirements, page 13-9
- Gateway Protocols, page 13-10
- Gateway Protocols and Core Feature Requirements, page 13-10
- Site-Specific Gateway Requirements, page 13-17

Core Feature Requirements

Gateways used in IP telephony applications must meet the following core feature requirements:

- Dual tone multifrequency (DTMF) relay capabilities

  DTMF relay capability, specifically out-of-band DTMF, separates DTMF digits from the voice stream and sends them as signaling indications through the gateway protocol (H.323, SCCP, MGCP, or SIP) signaling channel instead of as part of the voice stream or bearer traffic. Out-of-band DTMF is required when using a low bit-rate codec for voice compression because the potential exists for DTMF signal loss or distortion.

- Supplementary services support

  Supplementary services are typically basic telephony functions such as hold, transfer, and conferencing.

- Fax/modem support

  Fax over IP enables interoperability of traditional analog fax machines with IP telephony networks. The fax image is converted from an analog signal and is carried as digital data over the packet network. For more information, see Fax and Modem Support, page 13-19

- Unified CM redundancy support

  Cisco Unified Communications is based on a distributed model for high availability. Unified CM clusters provide for Unified CM redundancy. The gateways must support the ability to “re-home” to a secondary Unified CM in the event that a primary Unified CM fails. Redundancy differs from call survivability in the event of a Unified CM or network failure.

Refer to the gateway product documentation to verify that any IP Telephony gateway you select for an enterprise deployment can support the preceding core requirements. Additionally, every IP Telephony implementation has its own site-specific feature requirements, such as analog or digital access, DID, and capacity requirements (see Site-Specific Gateway Requirements, page 13-17).

Gateway Protocols

Cisco Unified CM (Release 3.1 and later) supports the following gateway protocols:

- H.323
- Media Gateway Control Protocol (MGCP)

Cisco Unified CM Release 4.0 and later supports Session Initiation Protocol (SIP) on the trunk side. The SIP trunk implementation has been enhanced in Cisco Unified CM releases 5.0 through 7.*x* to support more features.

Protocol selection depends on site-specific requirements and the installed base of equipment. For gateway configuration, MGCP might be preferred over H.323 or SIP due to simpler configuration. On the other hand, H.323 or SIP might be preferred over MGCP because of the robustness of the interfaces supported.

Simplified Message Desk Interface (SMDI) is a standard for integrating voice mail systems to PBXs or Centrex systems. Connecting to a voice mail system via SMDI and using either analog FXS or digital T1 PRI would require either SCCP or MGCP protocol because H.323 or SIP devices do not identify the specific line being used from a group of ports. Use of H.323 or SIP gateways for this purpose means the Cisco Message Interface cannot correctly correlate the SMDI information with the actual port or channel being used for an incoming call.

In addition, the Unified CM deployment model being used can influence gateway protocol selection. (Refer to the chapter on Unified Communications Deployment Models, page 5-1.)

Note Prior to deployment, check the Cisco IOS software release notes to confirm feature or interface support.

Gateway Protocols and Core Feature Requirements

This section describes how each protocol (SCCP, H.323, MGCP, and SIP) supports the following gateway feature requirements:

- DTMF Relay, page 13-10
- Supplementary Services, page 13-12
- Unified CM Redundancy, page 13-15

DTMF Relay

Dual-Tone Multifrequency (DTMF) is a signaling method that uses specific pairs of frequencies within the voice band for signals. A 64 kbps pulse code modulation (PCM) voice channel can carry these signals without difficulty. However, when using a low bite-rate codec for voice compression, the potential exists for DTMF signal loss or distortion. An out-of-band signaling method for carrying DTMF tones across a Voice over IP (VoIP) infrastructure provides an elegant solution for these codec-induced symptoms.

SCCP Gateways

The Cisco VG248 carries DTMF signals out-of-band using Transmission Control Protocol (TCP) port 2002. Out-of-band DTMF is the default gateway configuration mode for the VG248.

H.323 Gateways

The H.323 gateways, such as the Cisco 3800 series products, can communicate with Unified CM using the enhanced H.245 capability for exchanging DTMF signals out-of-band. The following is an example out-of-band DTMF configuration on a Cisco IOS gateway:

```
dial-peer voice 100 voip
destination-pattern 555….
session target ipv4:10.1.1.1
CODEC g729ar8
dtmf-relay h245-alphanumeric
preference 0
```

MGCP Gateway

The Cisco IOS-based platforms use MGCP for Unified CM communication. Within the MGCP protocol is the concept of *packages*. The MGCP gateway loads the DTMF package upon start-up. The MGCP gateway sends *symbols* over the control channel to represent any DTMF tones it receives. Unified CM then interprets these signals and passes on the DTMF signals, out-of-band, to the signaling endpoint. The global configuration command for DTMF relay is:

```
mgcp dtmf-relay VOIP codec all mode out-of-band
```

You must enter additional configuration parameters in the Unified CM MGCP gateway configuration interface.

DTMF relay is enabled by default and does not need additional configuration.

Note Use the **fm-package** command to enable DTMF via RFC 2833, available as of Cisco IOS Release 12.4(6)T.

SIP Gateway

The Cisco IOS-based platforms can use SIP for Unified CM communication. They support various methods for DTMF, but only the following methods can be used to communicate with Unified CM:

- Named Telephony Events (NTE), or RFC 2833
- Unsolicited SIP Notify (UN)
- Key Press Markup Language (KPML)

The following example shows a configuration for NTE:

```
dial-peer voice 100 voip
destination-pattern 555….
session target ipv4:10.1.1.1
session protocol sipv2
dtmf-relay rtp-nte
```

The following example shows a configuration for UN:

```
dial-peer voice 100 voip
destination-pattern 555….
session target ipv4:10.1.1.1
session protocol sipv2
dtmf-relay sip-notify
```

For more details on DTMF method selection, see the chapter on Media Resources, page 17-1.

Supplementary Services

Supplementary services provide user functions such as hold, transfer, and conferencing. These are considered fundamental requirements of any voice installation. Each gateway evaluated for use in an IP telephony network should provide support for supplementary services natively, without the use of a software media termination point (MTP).

SCCP Gateways

The Cisco SCCP gateways provide full supplementary service support. They also support FXS SCCP ports with Cisco IOS Release 12.4.9T. The SCCP gateways use the Gateway-to-Unified CM signaling channel and SCCP to exchange call control parameters.

H.323 Gateways

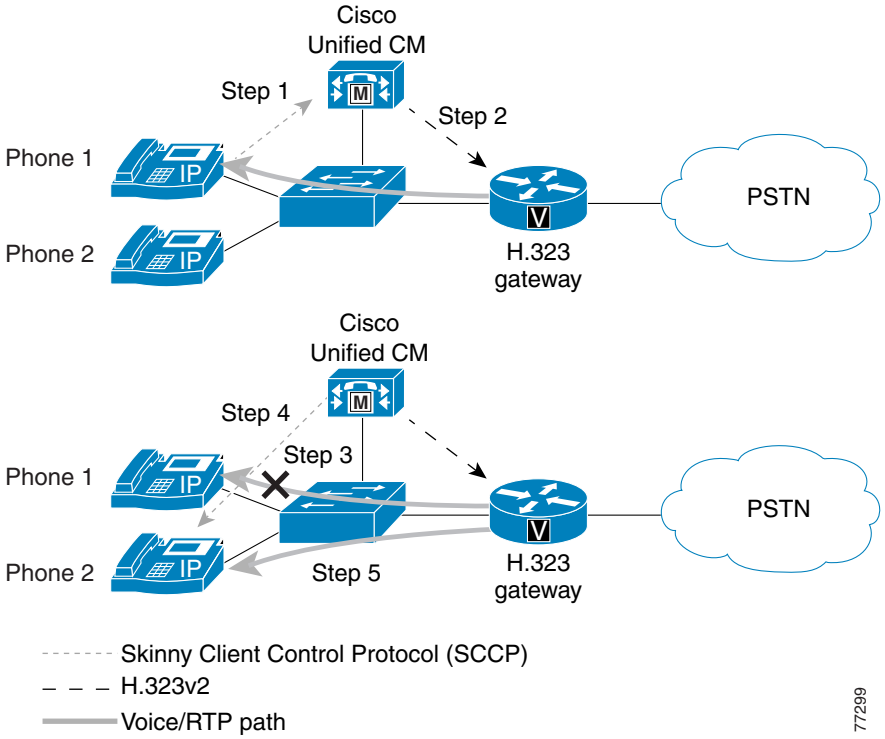
H.323v2 implements Open/Close LogicalChannel and the emptyCapabilitySet features. The use of H.323v2 by H.323 gateways, beginning in Cisco IOS Release 12.0(7)T and Cisco Unified CM Release 3.0 and later, eliminates the requirement for an MTP to provide supplementary services. With Unified CM Release 3.1 and later, a transcoder is allocated dynamically only if required during a call to provide access to G.711-only devices while still maintaining a G.729 stream across the WAN. Full support for H.323v2 is available in Cisco IOS Release 12.1.1T.

Once an H.323v2 call is set up between a Cisco IOS gateway and an IP phone, using the Unified CM as an H.323 proxy, the IP phone can request to modify the bearer connection. Because the Real-Time Transport Protocol (RTP) stream is directly connected to the IP phone from the Cisco IOS gateway, a supported voice codec can be negotiated.

Figure 13-1 and the following steps illustrate a call transfer between two IP phones:

1. If IP Phone 1 wishes to transfer the call from the Cisco IOS gateway to Phone 2, it issues a transfer request to Unified CM using SCCP.
2. Unified CM translates this request into an H.323v2 CloseLogicalChannel request to the Cisco IOS gateway for the appropriate SessionID.
3. The Cisco IOS gateway closes the RTP channel to Phone 1.
4. Unified CM issues a request to Phone 2, using SCCP, to set up an RTP connection to the Cisco IOS gateway. At the same time, Unified CM issues an OpenLogicalChannel request to the Cisco IOS gateway with the new destination parameters, but using the same SessionID.
5. After the Cisco IOS gateway acknowledges the request, an RTP voice bearer channel is established between Phone 2 and the Cisco IOS gateway.

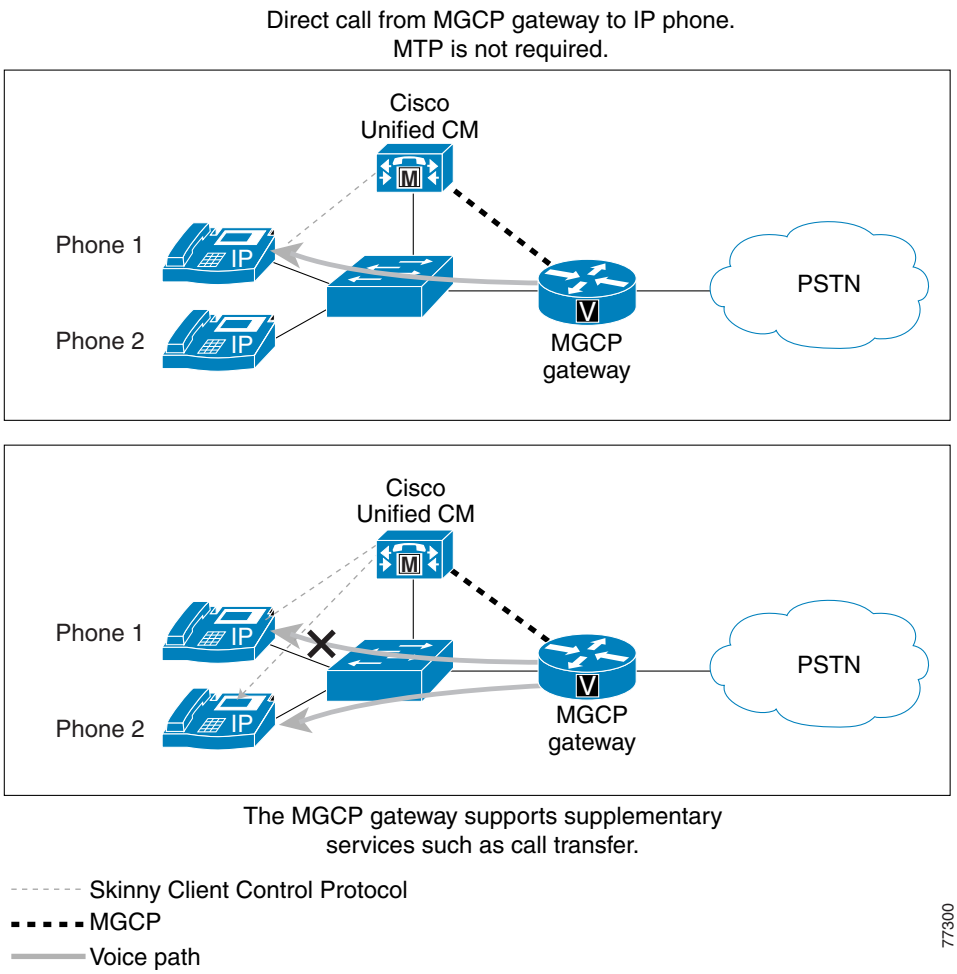
Figure 13-1 H.323 Gateway Supplementary Service Support



MGCP Gateway

The MGCP gateways provide full support for the hold, transfer, and conference features through the MGCP protocol. Because MGCP is a master/slave protocol with Unified CM controlling all session intelligence, Unified CM can easily manipulate MGCP gateway voice connections. If an IP telephony endpoint (for example, an IP phone) needs to modify the session (for example, transfer the call to another endpoint), the endpoint would notify Unified CM using SCCP. Unified CM then informs the MGCP gateway, using the MGCP User Datagram Protocol (UDP) control connection, to terminate the current RTP stream associated with the Session ID and to start a new media session with the new endpoint information. Figure 13-2 illustrates the protocols exchanged between the MGCP gateway, endpoints, and Unified CM.

Figure 13-2 MGCP Gateway Supplementary Service Support



SIP Gateway

The Unified CM SIP trunk interface to Cisco IOS SIP gateways supports supplementary services such as hold, blind transfer, and attended transfer. The support for supplementary services is achieved via SIP methods such as INVITE and REFER. For more details, refer to the following documentation:

- *Cisco Unified Communications Manager System Guide*, available at
  http://www.cisco.com/en/US/products/sw/voicesw/ps556/prod_maintenance_guides_list.html
- *Cisco IOS SIP Configuration Guide*, available at
  http://www.cisco.com/en/US/docs/ios/voice/sip/configuration/guide/12_4t/sip_12_4t_book.html

Unified CM Redundancy

An integral piece of the IP telephony architecture is the provisioning of low-cost, distributed PC-based systems to replace expensive and proprietary legacy PBX systems. This distributed design lends itself to the robust fault tolerant architecture of clustered Unified CMs. Even in its most simplistic form (a two-system cluster), a secondary Unified CM should be able to pick up control of all gateways initially managed by the primary Unified CM.

SCCP Gateways

Upon boot-up, the Cisco VG224, VG248, and ATA 188 gateways are provisioned with Unified CM server information. When these gateways initialize, a list of Unified CMs is downloaded to the gateways. This list is prioritized into a primary Unified CM and secondary Unified CM. In the event that the primary Unified CM becomes unreachable, the gateway registers with the secondary Unified CM.

H.323 VoIP Call Preservation for WAN Link Failures

H.323 VoIP call preservation enhancements for WAN link failures sustain connectivity for H.323 topologies where signaling is handled by an entity that is different from the other endpoint, such as a gatekeeper that provides routed signaling or a call agent (such as the Cisco BTS 10200 Softswitch, Cisco PGW2200 Softswitch, or Cisco Unified CM) that brokers signaling between the two connected parties. Call preservation is useful when a gateway and the other endpoint (typically a Cisco Unified IP Phone) are located at the same site but the call agent is remote and therefore more likely to experience connectivity failures.

H.323 call preservation covers the following types of failures and connections.

Failure Types:

- WAN failures that include WAN links flapping or degraded WAN links.
- Cisco Unified CM software failure, such as when the ccm.exe service crashes on a Unified CM server.
- LAN connectivity failure, except when a failure occurs at the local branch.

Connection Types:

- Calls between two Cisco Unified CM controlled endpoints under the following conditions:
  - During Unified CM reloads.
  - When a Transmission Control Protocol (TCP) connection used for signaling H.225.0 or H.245 messages between one or both endpoints and Unified CM is lost or flapping.
  - Between endpoints that are registered to different Unified CMs in a cluster, and the TCP connection between the two Unified CMs is lost.
  - Between IP phones and the PSTN at the same site.
- Calls between a Cisco IOS gateway and an endpoint controlled by a softswitch, where the signaling (H.225.0, H.245 or both) flows between the gateway and the softswitch and media flows between the gateway and the endpoint:
  - When the softswitch reloads.
  - When the H.225.0 or H.245 TCP connection between the gateway and the softswitch is lost, and the softswitch does not clear the call on the endpoint.
  - When the H.225.0 or H.245 TCP connection between softswitch and the endpoint is lost, and the softswitch does not clear the call on the gateway.

- Call flows involving a Cisco Unified Border Element (formerly, Cisco Multiservice IP-to-IP Gateway) running in media flow-around mode that reload or lose connection with the rest of the network.

Note that, after the media is preserved, the call is torn down later when either one of the parties hangs up or media inactivity is detected. In cases where there is a machine-generated media stream, such as music streaming from a media server, the media inactivity detection will not work and the call might hang. Cisco Unified CM addresses such conditions by indicating to the gateway that such calls should not be preserved, but third-party devices or the Cisco Unified Border Element would not do this.

Flapping is defined for this feature as the repeated and temporary loss of IP connectivity, which can be caused by WAN or LAN failures. H.323 VoIP calls between a Cisco IOS gateway and Cisco Unified CM may be torn down when flapping occurs. When Unified CM detects that the TCP connection is lost, it clears the call and closes the TCP sockets used for the call by sending a TCP FIN, without sending an H.225.0 Release Complete or H.245 End Session message. This is called *quiet clearing*. The TCP FIN sent from Unified CM could reach the gateway if the network comes up for a short duration, and the gateway will tear down the call. Even if the TCP FIN does not reach the gateway, the TCP keepalives sent from the gateway could reach Unified CM when the network comes up. Unified CM will send TCP RST messages in response to the keepalives because it has already closed the TCP connection. The gateway will tear down H.323 calls if it receives the RST message.

Configuration of H.323 VoIP call preservation enhancements for WAN link failures involves configuring the **call preserve** command. If you are using Cisco Unified Communications Manager, you must enable the Allow Peer to Preserve H.323 Calls parameter from the Service Parameters window.

The **call preserve** command causes the gateway to ignore socket closure or socket errors on H.225.0 or H.245 connections for active calls, thus allowing the socket to be closed without tearing down calls using those connections.

Example of H.323 VoIP Call Preservation for All Calls

The following configuration example enables H.323 VoIP call preservation for all calls:

```
voice service voip
 h323
  call preserve
```

MGCP Gateway

MGCP gateways also have the ability to fail over to a secondary Unified CM in the event of communication loss with the primary Unified CM. When the failover occurs, active calls are preserved.

Within the MGCP gateway configuration file, the primary Unified CM is identified using the **call-agent <hostname>** command, and a list of secondary Unified CM is added using the **ccm-manager redundant-host** command. Keepalives with the primary Unified CM are through the MGCP application-level keepalive mechanism, whereby the MGCP gateway sends an empty MGCP notify (NTFY) message to Unified CM and waits for an acknowledgement. Keepalive with the backup Unified CMs is through the TCP keepalive mechanism.

If the primary Unified CM becomes available at a later time, the MGCP gateway can “re-home,” or switch back to the original Unified CM. This re-homing can occur either immediately, after a configurable amount of time, or only when all connected sessions have been released. This is enabled through the following global configuration commands:

```
ccm-manager redundant-host <hostname1 | ipaddress1 > <hostname2 | ipaddress2>
[no] call-manager redundancy switchback [immediate|graceful|delay <delay_time>]
```

SIP Gateway

Redundancy with Cisco IOS SIP gateways can be achieved similarly to H.323. If the SIP gateway cannot establish a connection to the primary Unified CM, it tries a second Unified CM defined under another dial-peer statement with a higher preference.

By default the Cisco IOS SIP gateway transmits the SIP INVITE request 6 times to the Unified CM IP address configured under the dial-peer. If the SIP gateway does not receive a response from that Unified CM, it will try to contact the Unified CM configured under the other dial-peer with a higher preference.

Cisco IOS SIP gateways wait for the SIP 100 response to an INVITE for a period of 500 ms. By default, it can take up to 3 seconds for the Cisco IOS SIP gateway to reach the backup Unified CM. You can change the SIP INVITE retry attempts under the **sip-ua** configuration by using the command **retry invite** <*number*>. You can also change the period that the Cisco IOS SIP gateway waits for a SIP 100 response to a SIP INVITE request by using the command **timers trying** <*time*> under the **sip-ua** configuration.

One other way to speed up the failover to the backup Unified CM is to configure the command **monitor probe icmp-ping** under the **dial-peer** statement. If Unified CM does not respond to an Internet Control Message Protocol (ICMP) echo message (ping), the dial-peer will be shut down. This command is useful only when the Unified CM is not reachable. ICMP echo messages are sent every 10 seconds.

The following commands enable you to configure Unified CM redundancy on a Cisco IOS SIP gateway:

```
sip-ua
 retry invite <number>
 timers trying <time>

dial-peer voice 101 voip
 destination-pattern 2...
 session target ipv4:10.1.1.101
 preference 0
 monitor probe icmp-ping
 session protocol sipv2


dial-peer voice 102 voip
 destination-pattern 2...
 session target ipv4:10.1.1.102
 preference 1
 monitor probe icmp-ping
 session protocol sipv2
```

Site-Specific Gateway Requirements

Each IP Telephony implementation has its own site-specific requirements. The following questions can help you with IP Telephony gateway selection:

- Is the PSTN (or PBX) access analog or digital?
- What type of analog (FXO, FXS, E&M, DID, CAMA) or digital (T1, E1, CAS, CCS) interface is required for the PSTN or PBX?
- If the PSTN access is digital, what type of signaling is required (T1 CAS, Q.931 PRI, E1 CAS, or R2)?

- What type of signaling does the PBX currently use?
  - FXO or FXS: loop start or ground start
  - E&M: wink-start, delay-start, or immediate-start
  - E&M: type I, II, III, IV, or V
  - T1: CAS, Q.931 PRI (User-Side or Network-Side), QSIG, DPNSS, or Proprietary d-channel (CCS) signaling
  - E1: CAS, R2, Q.931 PRI (User-Side or Network-Side), QSIG, DPNSS, Proprietary d-channel (CCS) signaling
- What type of framing (SF, ESF, or G.704) and line encoding (B8ZS, AMI, CRC-4, or HDB3) does the PBX currently use?
- Does the PBX require passing proprietary signaling? If so, which time slot is the signaling passed on, and is it HDLC-framed?
- What is the required capacity of the gateway; that is, how many channels are required? (Typically, if 12 or more voice channels are required, then digital is more cost effective than an analog solution.)
- Is Direct Inward Dialing (DID) required? If so, specify analog or digital.
- Is Calling Line ID (CLID) needed?
- Is Calling Name needed?
- What types of fax and modem support are required?
- What types of voice compression are required?
- What types of supplementary services are required?
- Will the PBX provide clocking, or will it expect the Cisco gateway to provide clocking?
- Is rack space available for all needed gateways, routers, and switches?

Note Direct Inward Dial (DID) refers to a private branch exchange (PBX) or Centrex feature that permits external calls to be placed directly to a station line without use of an operator.

Note Calling Line Identification (CLI, CLID, or ANI) refers to a service available on digital phone networks to display the calling number to the called party. The central office equipment identifies the phone number of the caller, enabling information about the caller to be sent along with the call itself. CLID is synonymous with Automatic Number Identification (ANI).

Cisco Unified Communications gateways are able to inter-operate with most major PBX vendors, and they are EIA/TIA-464B compliant.

The site-specific and core gateway requirements are a good start to help narrow the possible choices. Once you have defined the required features, you can make a gateway selection for each of the pertinent configurations, whether they are single-site enterprise deployments of various sizes and complexities or multisite enterprise deployments.

The following tables summarize the features and interface types supported by the various Cisco gateway models.

Note In the following tables, the Cisco IOS and Unified CM release numbers refer to the minimum release that can support the listed feature on a particular gateway platform. For more information about Cisco IOS features, refer to the Cisco Feature Navigator located at http://tools.cisco.com/ITDIT/CFN/jsp/index.jsp.

Fax and Modem Support

This section describes the fax and modem support available with Unified CM and Cisco voice gateways. This section first presents brief overviews of fax and modem support on Cisco voice gateways, followed by a listing of supported platforms and example configuration files.

Gateway Support for Fax Passthrough and Fax Relay

Fax over IP enables interoperability of traditional analog fax machines with IP Telephony networks. The fax image is converted from an analog signal and is carried digitally over the packet network.

In its original form, fax data is digital and is contained in High-Level Data Link Control (HDLC) frames. However, to transmit across a traditional PSTN, these digital HDLC frames are modulated onto an analog carrier. While this analog carrier is necessary for effective faxing in PSTN environments, it is not ideal for the type of digital transport used by IP packet networks. Therefore, specific transport methods have been devised for successful transport of fax transmissions over an IP infrastructure.

The two main methods for transporting fax over IP are passthrough and relay. Passthrough is the simplest method, and it works by sampling and digitizing the analog fax signal just like a voice codec does for human speech. While there are a number of codecs available, passthrough on Cisco voice gateways always uses the G.711 codec for carrying fax information because it offers the least distortion of the analog fax signals. If a high-compression codec is being used by the original voice call, then passthrough uses an upspeed feature to change the codec to G.711. Passthrough is also commonly referred to as Voice Band Data (VBD), and Cisco provides two versions of passthrough: modem passthrough and fax pass-through. The names of these two passthrough versions are derived from how they are configured in the Cisco IOS command line interface (CLI). Additional differences between these passthrough versions center around their switchovers and triggering tones, and these are discussed in more detail in the following paragraphs.

Modem passthrough typically uses Cisco proprietary Named Signaling Event (NSE) packets to switch the call from voice mode to passthrough mode in what is commonly termed an NSE-based switchover. This switchover from voice mode to passthrough is an important concept for fax pass-through and relay as well. Every call on a Cisco voice gateway starts as a voice call, and the proper switchover occurs only when the gateway determines that the call is truly a fax call.

The modem passthrough feature is triggered by a 2100 Hz CED or ANSam tone at the beginning of a fax or modem call. The CED tone is associated with G3 faxes and low-speed modems, while the ANSam tone is used by SG3 faxes and high-speed modems. Historically, when the ANSam or CED tone was detected, modem passthrough used Cisco proprietary NSE packets to signal the remote voice gateway of the switchover from voice mode to modem passthrough. Now, however, in addition to an NSE-based switchover, modem passthrough also supports a protocol-based switchover using the H.323 or SIP call control protocols. When modem passthrough is configured to handle the switchover using H.323 or SIP, it also will use a standards-based NTE message to optionally signal the remote voice gateway to disable its echo cancellers. These enhancements to modem passthrough allow for increased interoperability with third-party devices and are found in Cisco IOS Release 12.4(24)T and higher.

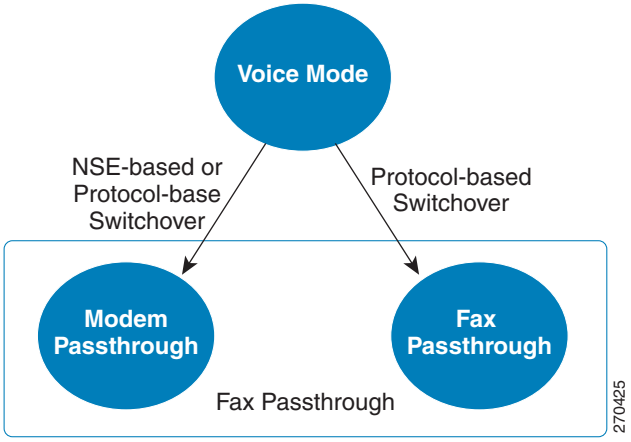
Despite its name, modem passthrough is also widely used for fax calls. You can activate it in the Cisco IOS command line interface (CLI) by using the **modem passthrough** command for H.323, SIP, and SCCP voice gateways or the **mgcp modem passthrough** command for MGCP voice gateways.

Fax pass-through does not support an NSE-based switchover as modem passthrough does. Instead, it always relies on the underlying call control protocol to switch the call from voice mode to fax pass-through. Fax pass-through supports only a protocol-based switchover using the call control protocols of H.323 and SIP. Because fax pass-through utilizes the call control protocol for its switchover, it will typically interoperate with third-party devices.

Fax pass-though is triggered by the detection of V.21 flags associated with G3 fax calls. Therefore, this transport method does not work for modems or SG3 fax calls. The command to enable fax pass-through on H.323 and SIP voice gateways is **fax protocol pass-through**.

Figure 13-3 highlights the two different passthrough implementations employed by Cisco voice gateways for fax calls.

Figure 13-3 Cisco Passthrough Implementations for Fax Calls

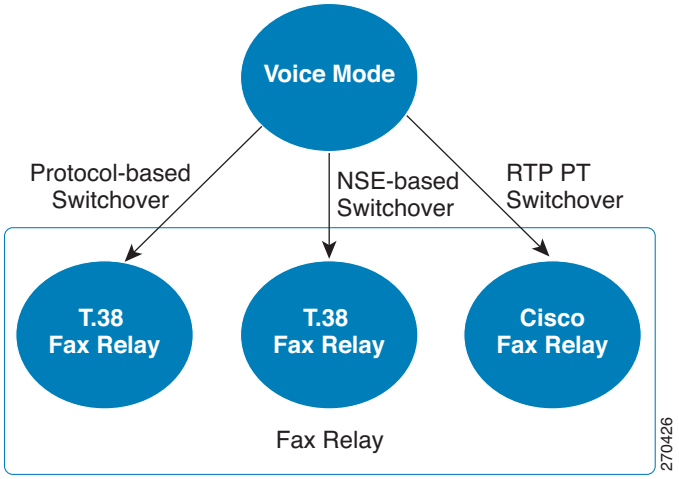


Relay is the other main method for transporting fax over IP, and its implementation is a bit more complicated than passthrough. Relay strips off the analog carrier from the fax signal in a process known as *demodulation* to expose the fax HDLC data frames. The pertinent information in these HDLC frames is then removed and efficiently packaged in a fax relay protocol to be transported to the gateway on the other side. When received on the other side, the fax information is pulled from the relay protocol, reconstructed back into fax HDLC frames, and modulated onto an analog carrier for transmission to a fax machine.

Cisco supports two versions of fax relay, T.38 and Cisco fax relay. An ITU standard, T.38 allows Cisco gateways to interoperate with third-party devices that also support the T.38 specification. In most scenarios, T.38 fax relay uses the call control protocol to switch from voice mode to T.38 fax relay mode, and this is referred to as protocol-based or standards-based T.38 fax relay. However, it is also possible to configure T.38 fax relay to switch over using Cisco proprietary NSEs in what is termed NSE-based T.38 fax relay. To ensure third-party interoperability, protocol-based T.38 must be utilized.

Cisco fax relay is a pre-standard implementation, and it is proprietary to Cisco voice gateways. It is also the default fax transport configuration on nearly all Cisco voice gateways. Unlike the NSE or protocol-based methods used by T.38 fax relay and passthrough, Cisco fax relay transitions from voice to relay mode utilizing specific RTP dynamic payload types (PT). Figure 13-4 illustrates the Cisco fax relay methods.

Figure 13-4 Cisco Relay Implementations for Fax Calls



Fax relay mode, and more specifically T.38, is the preferred method to transport fax traffic. However, if T.38 fax relay is not supported, then Cisco fax relay or passthrough can be used as an alternative.

Best Practices

The following recommendations and guidelines can assist you in best implementing fax support on Cisco voice gateways:

- When using QoS, make every effort to minimize the following:
  - Packet loss
  - Delay
  - Delay variation (jitter)

  All transmissions of fax over IP are extremely sensitive to packet loss. Even minimal packet loss can cause fax failures. If packet loss is a problem in your network, then you should use the redundancy feature in T.38 fax relay. Also, ensure that constant packet delay on the network does not exceed 1 second and that delay variation (jitter) does not exceed 300 milliseconds for T.38 and Cisco fax relay. When passthrough is used, the jitter should follow VoIP design best practices and not exceed 30 ms. For detailed information about implementing QoS in a Cisco Unified Communications network, refer to the *Enterprise QoS Solution Reference Network Design Guide*, available at

  http://www.cisco.com/go/designzone
- The following tips can help ensure the integrity of the fax calls:
  - Use call admission control (CAC) to ensure that calls are not admitted if they exceed the specified total bandwidth limit. The following table lists approximate fax call bandwidth usage for the common fax transport methods.

Fax Transport Method	Bandwidth[1]
Modem Passthrough or Fax Pass-through (G.711)	83 kbps
Modem Passthrough with Redundancy	170 kbps
T.38 (no redundancy)	25 kbps

Fax Transport Method	Bandwidth[1]
T.38 (high-speed redundancy level set to 1)	41 kbps
T.38 (high-speed redundancy level set to 2)	57 kbps
Cisco Fax Relay	48 kbps

1. Bandwidth values are approximate with Ethernet or Frame Relay L2 headers. T.38 and Cisco fax relay bandwidth values are peak and occur only during the sending of a fax page at 14.4 kbps.

  - Disable call waiting on all dedicated modem and fax ports.

- T.38 fax relay provides the best fax performance based on network considerations and is the recommended transport method for fax traffic.

  To insure interoperability with other vendor's T.38 products, use protocol-based T.38.

  NSE-based T.38 must be used for communicating with certain Cisco voice gateways, such as the Cisco VG248 and any Cisco IOS SCCP gateways. For older versions of Unified CM with limited support for protocol-based T.38, NSE-based T.38 fax relay is a valid alternative.

  In Unified CM scenarios where T.38 is to be deployed among gateways running a variety of call signaling protocols, protocol-based T.38 should be the first choice. The latest release of Cisco Unified CM supports protocol-based T.38 with H.323, SIP, and MGCP call control protocols. If protocol-based T.38 is not supported in your installed version of Cisco Unified CM or if SCCP gateways are involved, then NSE-based T.38 should be used. To verify if your version of Unified CM supports protocol-based T.38, refer to the Cisco Unified Communications Manager release notes available at

  http://www.cisco.com/en/US/products/sw/voicesw/ps556/prod_release_notes_list.html

- T.38 fax relay is supported on most of the current Cisco voice gateways, especially those running Cisco IOS. For details, refer to the product data sheets for your specific gateway models.
- Error Correction Mode (ECM) is a negotiated feature on fax calls, and it ensures that fax pages are received error-free. However, in its effort to retransmit any errors, ECM can lead to increased fax transmission times and call failures. If desired, you can disable ECM on the gateway itself rather than disabling it on multiple fax machines. However, if packet drops or other IP or PSTN impairments occur, the fax image quality might deteriorate. Therefore, you should disable ECM only after considering whether you want to risk compromising image quality rather than experiencing longer call durations or dropped calls. You should also monitor and evaluate the network to identify and resolve the impairments that are causing the fax page errors.

Super Group 3 Fax Support

Commonly referred to as "high-speed" fax or V.34 fax, the Super Group 3 (SG3) classification uses V.34 modulation to increase the maximum fax page transmission speed to 33.6 kbps. SG3 fax machines are backward compatible with standard G3 fax machines that support a maximum page transmission speed of 14.4 kbps.

Cisco IOS gateways with Cisco IOS Release 12.4.4T and later offer support for Super Group 3 (SG3) fax transmissions when T.38 or Cisco fax relay are configured; however, only Group 3 speeds are negotiated. For more information on this feature, refer to *Cisco IOS Fax, Modem, and Text Support over IP Configuration Guide, Cisco IOS Release 15.1M&T*, available at

http://www.cisco.com/en/US/docs/ios/voice/fax/configuration/guide/15_1/vf_15_1_book.html

If it is necessary to transport SG3 high-speed faxes at their native speeds, then modem passthrough must be used. With the release of Cisco IOS version 15.1.1T, a new feature will provide native support for SG3 faxes by T.38 fax relay.

Gateway Support for Modem Passthrough and Modem Relay

In general, there are three mechanisms for supporting modem sessions over an IP network using voice gateways:

- Modem passthrough
- Cisco Modem Relay
- Secure Modem Relay (Secure Communication Between STE Endpoints)

Each of these mechanisms can transport modem calls, but the relay methods are restrictive in that only certain modem modulations are supported. Modem passthrough, on the other hand, can usually handle any modulation.

An important concept to understand when dealing with the transport of modem signals across IP networks is the switchover that must occur on the gateway. Every call on a Cisco gateway begins as a voice call initially. Even if the call is between modems, the call will be set up as a voice call first. Then, once the gateway is sure that the call is truly a modem call, a switchover occurs that converts the gateway from voice call mode to a modem passthrough or modem relay mode. There are various switchover methods to transition a call from voice mode to modem passthrough or relay.

As discussed previously in the section on Gateway Support for Fax Passthrough and Fax Relay, page 13-19, modem passthrough uses proprietary NSE packets or the H.323/SIP protocol stack to switch a voice call into passthrough mode. When modem signals are detected, the gateways can use these NSE messages to inform each other of the impending modem call. Special messages within the H.323 or SIP call control protocols can also be used. The gateways then make adjustments to better handle the transport of the modem signals. These adjustments include up-speeding the voice codec to G.711, disabling Voice Activity Detection (VAD), and disabling the echo cancellers if necessary. Because modem passthrough simply samples the analog modem signal using the G.711 codec, it should handle any modem modulation, but not always at the highest speeds.

Cisco Modem Relay is a proprietary implementation that efficiently transports V.34 modem calls over an IP network. V.90 calls are also supported, but they are forced to train down to V.34 speeds. As with modem passthrough, NSE packets are used to handle the switchover to Cisco Modem Relay from voice mode.

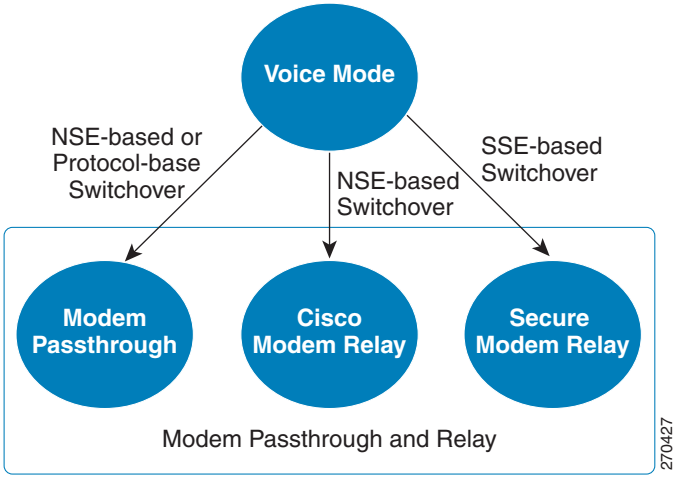
Secure modem relay, which is also referred to as Secure Communication Between STE Endpoints, allows for the transport of secure telephone calls over an IP infrastructure. Special devices known as Secure Terminal Equipment (STE) transmit encrypted voice using the V.32 modulation. Secure modem relay is designed to handle the transport of information between STEs in Unified CM environments with SCCP and MGCP gateways. Secure modem relay is not compatible with Cisco Modem Relay. One of the main reasons is that the switchover for secure modem relay does not use NSEs but instead uses V.150.1-based State Signaling Event (SSE) messages.

Secure modem relay is designed specifically for transporting STE signals and is almost never used outside of government or defense-related deployments. In most cases, Cisco Modem Relay or modem passthrough should be used for transporting modem calls. For more information on secure modem relay, refer to *Secure Communication Between IP-STE Endpoint and Line-Side STE Endpoint*, available at

http://www.cisco.com/en/US/docs/ios/12_4t/12_4t4/htv1501.html

Figure 13-5 summarizes the Cisco modem transport implementations. Modem relay should be used whenever possible because it offers the most bandwidth efficiency and tolerance for network impairments when compared to modem passthrough. The disadvantage of modem relay is that it is quite restrictive on the modulations supported, while modem passthrough can handle any modem modulation.

Figure 13-5 Cisco Passthrough and Relay Implementations for Modem Calls



Best Practices

Observe the following recommended best practices to ensure optimum performance of modem traffic transported over an IP infrastructure:

- Ensure that the IP network is enable for Quality of Service (QoS) and that you adhere to all of the recommendations for providing QoS in the LAN, MAN, and WAN environments. Every effort should be made to minimize the following parameters:
  - Packet loss — Fax and modem traffic requires an essentially loss-free transport. A single lost packet can result in retransmissions.
  - Delay
  - Delay variation (jitter)

  For more information, refer to the *Enterprise QoS Solution Reference Network Design Guide*, available at

  http://www.cisco.com/go/designzone
- Use call admission control (CAC) to ensure that calls are not admitted if they exceed the specified total bandwidth limit. For planning purposes, assume that modem passthrough calls will always consume approximately 83 kbps of bandwidth, or 170 kbps with redundancy enabled, regardless of the modem modulation being transported. Modem relay bandwidth is sporadic because of the nature of modem communications, but plan for peaks of about 45 kbps for the maximum V.34 connection speed of 33.6 kbps. The bandwidth values cited here are approximations and assume Ethernet or frame relay as the L2 transport.
- Use modem relay whenever possible. Modem passthrough should be used for any modulations not supported by modem relay.

- Do not use the IP network to connect modems that will be used to troubleshoot or diagnose problems with the IP network. In this case, the modems used to troubleshoot the devices that compose the IP infrastructure should be connected to a plain old telephone service (POTS).
- Because of the NSE switchover utilized by Cisco modem relay and modem passthrough, gateways using different call control protocols can easily communicate with one another. For example, an MGCP gateway and an H.323 gateway connected to Unified CM can successfully negotiate Cisco modem relay or modem passthrough because the NSE switchover occurs within the RTP voice media stream that has already been set up by Unified CM.
- Disable call waiting on all dedicated modem and fax ports.

V.90 Support

Currently, Cisco equipment supports only V.34 modems. Although V.90 modems will function on existing hardware, and speeds higher than V.34 speeds can be achieved, full V.90 support cannot be guaranteed.

Supported Platforms and Features

The following Cisco platforms support fax and modem features:

- Cisco IOS Gateways support:
  - Modem passthrough
  - Fax passthrough for the H.323 and SIP protocols
  - T.38 fax relay. Both NSE and protocol-based switchovers for T.38 are supported, except in the case of SCCP where only NSE-based T.38 fax relay is supported.
  - Cisco fax relay. The Cisco AS5350, AS5400, and AS5850 using Nextport DSP cards do not support Cisco fax relay. The PVDM3 DSP modules also do not support Cisco fax relay.
  - Cisco modem relay
- Cisco non-IOS gateways:
  - The Cisco VG248 supports modem passthrough, NSE-based T.38 fax relay, and Cisco fax relay.
  - The Cisco 6608 and 6624 support only modem passthrough and Cisco fax relay.
  - The Cisco ATAs support modem passthrough for fax calls only. Using modem passthrough with an ATA for modem calls is not officially supported.

Note The fax and modem support information presented here is valid beginning with Cisco IOS Release 12.4(9)T for the Cisco IOS gateways and Release 1.3.1 of the Cisco VG248 Analog Phone Gateway.

Platform Protocol Support

Common call control protocols used today in enterprise solutions include H.323, Session Initiation Protocol (SIP), Media Gateway Control Protocol (MGCP), and Skinny Client Control Protocol (SCCP). Not all Cisco voice platforms support all of these protocols or all of the fax and modem features, thus raising interoperability issues. Additional interoperability issues occur when mixing Cisco IOS

gateways such as the Cisco 2800 Series or the Cisco 3800 Series with non-IOS gateways such as the VG248. This section lists the combinations of gateways that provide support for interoperability of fax, modem, and protocol features.

Some of the common combinations of protocols in a network include: MGCP and H.323; SCCP and H.323; SCCP and SIP; MGCP and SIP; H.323 and SIP; and SCCP and MGCP.

Table 13-3 lists the protocol combinations that currently support fax and modem interoperability.

Table 13-3 Fax and Modem Features Supported with Various Combinations of Call Control Protocols

Protocol Combinations	Modem Relay	Modem Passthrough[1]	T.38 Fax Relay	Cisco Fax Relay	Fax Passthrough
Unified CM using MGCP combined with Unified CM using H.323 or SIP	Yes	Yes	Yes[2]	Yes	No
Unified CM using MGCP combined with Unified CM using MGCP	Yes	Yes	Yes[2]	Yes	No
SCCP combined with Unified CM using H.323 or SIP	Yes	Yes	Yes[3]	Yes	No
SCCP combined with Unified CM using MGCP	Yes	Yes	Yes[3]	Yes	No
Unified CM using H.323 combined with H.323 or SIP	Yes	Yes	Yes[2]	Yes	Yes
Unified CM using SIP combined with H.323 or SIP	Yes	Yes	Yes[2]	Yes	Yes

1. Modem passthrough works for both modem and fax passthrough calls.
2. NSE-based T.38 fax relay works, but protocol-based T.38 fax relay depends on the version of Unified CM. For version information, refer to the Cisco Unified Communications Manager release notes available at http://www.cisco.com/en/US/products/sw/voicesw/ps556/prod_release_notes_list.html.
3. SCCP protocol works only with NSE-based T.38 fax relay.

Note Table 13-3 is a general reference. You should be aware that specific products might have limitations that are not listed in this table. For example, the Cisco ATA supports H.323, SIP, and SCCP call control protocols, but only modem passthrough is supported no matter which call control protocol is used.

Gateway Configuration Examples

For detailed configuration information about fax and modem support on Cisco gateways, refer to the *Cisco IOS Fax, Modem, and Text Support over IP Configuration Guide*, which is available at

http://www.cisco.com/en/US/docs/ios/voice/fax/configuration/guide/12_4t/vf_12_4t_book.html

Gateways for Video Telephony

Video gateways terminate video calls into an IP telephony network or the PSTN. Video gateways are different from voice gateways because they have to interact with the ISDN trunks that support video and convert that call to a video call on the IP network using protocols such as H.323 or SIP. Enterprises can consider separate gateways for voice calls and video calls, or they can have integrated gateways that route both voice and video calls.

The following key considerations can help you decide if you need separate gateways for voice and video or an integrated gateway:

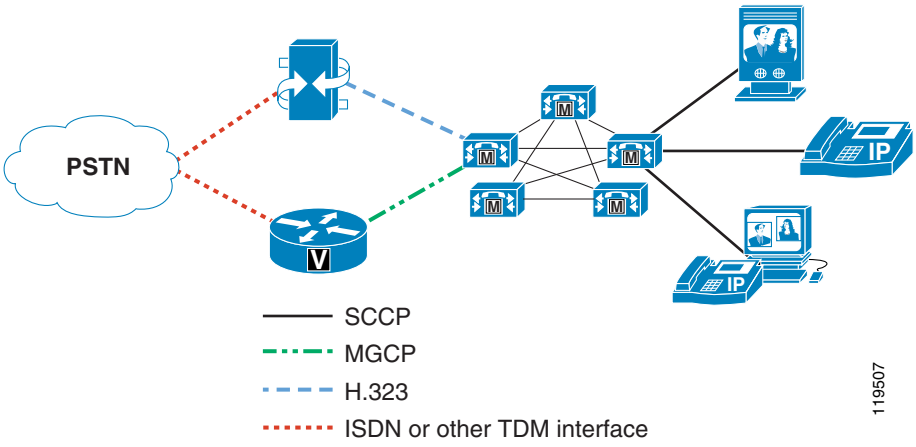
- Dial plan — If the enterprise has the flexibility of a separate dial plan for video users, it can use separate video gateways that allow it to keep existing enterprise dial plans.
- Video users — If the enterprise has a large number of users who primarily use voice rather than video, then Cisco recommends using separate video gateways to service the video call users.
- Locations — If the enterprise has a large number of distributed locations with video users at many locations, then Cisco recommends using an integrated gateway to reduce total cost of ownership (TCO).
- Additional video capabilities such as video IVR, auto attendant, and bonding across trunks — Dedicated video gateways might support advanced features that integrated gateways do not support.
- Protocol — Gateway protocol can be an important factor to align with enterprise policies and standards.
- Capacity — Dedicated gateways might support lower simultaneous call volumes, while integrated gateways should have higher capacity because they can support voice calls in addition to video.
- Device management — Ease of maintenance, management, and troubleshooting can be an important factor. Dedicated gateways provide a better user interface (GUI) for management and configuration, while integrated gateways can provide better troubleshooting. However, these factors are dependent on the respective products.

Dedicated Video Gateways

Enterprises that have an extensive voice infrastructure with voice gateways can add video gateways so that users can make video calls through them to the PSTN. The Cisco Unified Videoconferencing 3500 and 5200 Series Video Gateways can be used for that purpose.

Figure 13-6 shows an enterprise deployment that can use existing protocols for its voice gateways and add video gateways so that Unified CM users can make voice and video calls to the PSTN.

Figure 13-6 Unified CM System with Separate PSTN Lines for Voice and IP Video Telephony



The Unified Videoconferencing gateways, while excellent for video calls, do not support all of the features that Cisco Voice Gateways offer. The Unified Videoconferencing gateways have the following characteristics:

- They support only H.323 and H.320.
- They are standalone devices that cannot be integrated into Cisco IOS Routers or Cisco Catalyst Switches.
- They support only T1/E1-PRI and ISDN BRI.
- They support H.261, H.263, and H.264 video codecs
- They support only G.711, G.722, G.722.1, G.723.1 and G.728; they do not support G.729 audio.
- They support the H.245 Empty Capabilities Set (ECS).
- They support T.120 and H.239 data sharing protocol
- They support H.235 encryption
- They do not support many of the manageability and troubleshooting capabilities inherent in Cisco Voice Gateways.

As a result of these differences in the products, Cisco Unified Videoconferencing 3500 Series gateways are not recommended as replacements for Cisco Voice Gateways. IP Telephony customers who want to add video to their communications environment should deploy both types of gateways and use the Cisco Voice Gateways for all voice calls and use the Cisco Unified Videoconferencing 3500 Series gateways for video calls only. Customers might also have to procure separate circuits for voice and video from their PSTN service provider, depending on which model of Cisco IOS Gateway is deployed.

With separate voice and video gateways (see Figure 13-6), the route plans must also be separate for both inbound and outbound calls. For inbound calls, there is no way to have a single Direct Inward Dial (DID) extension for a user who wants to be able to receive both voice and video calls. Typically, each user will already have a DID for voice calls. When you introduce video into the scenario, users will have to be dialed some other way, such as via a second DID number or by dialing the main number of the video gateway and then entering the users video extension when prompted by the Interactive Voice Response (IVR). For outbound calls, there is no way to have a single PSTN access code for both voice and video calls. Typically, users will already have a well-known access code for voice (such as 9 in most US enterprises), but when you introduce video into the scenario, they will have to dial some other access code to place outbound video calls.

Another consideration for deploying two types of gateways is the placement of those gateways. Typically, enterprises have many PSTN gateway resources consolidated at their central site(s), and each branch office has some local gateway resources as well. For instance, Cisco Catalyst 6500 gateways may be deployed at the central site with several T1/E1 circuits connected to them, while Cisco Integrated Services Routers (ISRs) may be deployed at each branch office with either analog or digital trunks to the local CO. When video is introduced into this scenario, the customer must also determine the number of PSTN circuits they will need for video and where the video gateways will be placed. For instance, will they deploy only a few Cisco Unified Videoconferencing video gateways at the central site, or will they also deploy them at each branch office?

Finally, consider how calls will be routed across the IP network to a remote gateway for the purpose of providing toll bypass, and how calls will be re-routed over the PSTN in the event that the IP network is unavailable or does not have enough bandwidth to complete the call. More specifically, do you want to invoke automated alternate routing (AAR) for video calls?

Integrated Video Gateways

Enterprises may consider an integrated device for voice and video gateway functionality. This provides the enterprise the advantages of managing fewer devices and keeping the dial plan simple. The gateway processes the call as a voice call if it is voice and as a video call if it is video.

The Cisco IOS Integrated Video Gateway has the following characteristics:

- Supports Cisco ISO-13871 bonding
- Provides H.320, H.323, and SIP support
- Supports existing voice codec and H.254 video codec
- Provides extensive called and calling transformation capabilities
- Provides extensive logging and troubleshooting capabilities

The following considerations apply for deploying Cisco IOS Integrated Video gateways:

- Consider the capacity needed on PSTN links for additional video calls.
- Consider the need of devices to use data applications such as T.120 and the additional bandwidth that will be used on the IP network.
- Consider if users need features such far-end camera control or DTMF that is used for conferences that the H.320 gateway needs to support.

Routing Inbound Calls from the PSTN

Use one of the following methods to route inbound calls from the PSTN:

- Assign at least two different directory numbers to each video-enabled device in the Unified CM cluster, with one line for audio and another line for video. With this method, the outside (PSTN) caller must dial the correct number to enable video.
- For video calls, have outside callers dial the main number of the video gateway.
  Cisco Unified Videoconferencing gateways offer an integrated IVR that prompts the caller to enter the extension number of the party they are trying to reach. Unified CM will then recognize that it is

a video call when ringing the destination device. This method relieves the caller from having to remember two different DID numbers for each called party, but it adds an extra step to dialing an inbound video call.

Note The outside video endpoints must support DTMF in order to enter the extension of the called party at the IVR prompt.

The following example illustrates the second method:

A user has a Cisco Unified IP Phone 7960 attached to a PC running Cisco Unified Video Advantage. The extension of the IP Phone is 51212, and the fully qualified DID number is 1-408-555-1212. To reach the user from the PSTN for a voice-only call, people simply dial the DID number. The CO sends calls to that DID number through T1-PRI circuit(s) connected to a Cisco Voice Gateway. When the call is received by the gateway, Unified CM knows that the gateway is capable of audio only, so it negotiates only a single audio channel for that call. Conversely, for people to reach the user from the PSTN for a video call, they must dial the main number of the video gateway and then enter the user's extension. For example, they might dial 1-408-555-1000. The CO would send calls to that number through the T1-PRI circuit(s) connected to a Cisco Unified Videoconferencing 3500 Series video gateway. When the call is received by the gateway, an IVR prompt asks the caller to enter the extension of the person they are trying to reach. When the caller enters the extension via DTMF tones, Unified CM knows that the gateway is capable of video, so it negotiates both audio and video channels for that call.

Gateway Digit Manipulation

The Cisco Unified Videoconferencing 3500 Series Gateways cannot manipulate digits for calls received from the PSTN. It takes the exact number of digits passed to it in the Q.931 Called Party Number field and sends them all to Unified CM. Therefore, Unified CM must manipulate the digits in order to match the directory number (DN) of the destination device. For instance, if the circuit from the CO switch to the gateway is configured to pass 10 digits but the extension of the called party is only five digits, Unified CM must strip off the leading five digits before attempting to find a matching DN. You can implement this digit manipulation in one of the following ways:

- By configuring the Significant Digits field on the H.323 gateway device or on the H.225 gatekeeper-controlled trunk that carries the incoming calls from the Cisco Unified Videoconferencing gateway

  This method enables you to instruct Unified CM to pay attention to only the least-significant N digits of the called number. For example, setting the Significant Digits to 5 will cause Unified CM to ignore all but the last 5 digits of the called number. This is the easiest approach, but it affects all calls received from that gateway. Thus, if you have variable-length extension numbers, this is not the recommended approach.

- By configuring a translation pattern and placing it in the calling search space of the H.323 gateway device or of the H.225 gatekeeper-controlled trunk that carries the incoming calls from the Cisco Unified Videoconferencing gateway

  This method enables Unified CM to match calls to the full number of digits received, to modify the called number, and then to continue performing digit analysis on the resulting modified number. This approach is slightly more complex than the preceding method, but it is more flexible and enables you to use a finer granularity for matching calls and for specifying how they will be modified.

Routing Outbound Calls to the PSTN

Use one of the following methods to route outbound calls to the PSTN:

- Assign different access codes (that is, different route patterns) for voice and video calls. For example, when the user dials 9 followed by the PSTN telephone number they are trying to reach, it could match a route pattern that directs the call out a voice gateway. Similarly, the digit 8 could be used for the route pattern that directs calls out a video gateway.
- Assign at least two different directory numbers on each video-enabled device in the Unified CM cluster, with one line for audio and another line for video. The two lines can then be given different calling search spaces. When users dial the access code (9, for example) on the first line, it could be directed out a voice gateway, while dialing the same access code on the second line could direct the call out a video gateway. This method alleviates the need for users to remember two different access codes but requires them to press the correct line on their phones when placing calls.

Gateway Service Prefixes

The Cisco Unified Videoconferencing Gateways use service prefixes to define the speed for outbound calls. When you configure a service prefix in the gateway, you must choose one of the following speeds:

- Voice-only
- 128 kbps
- 256 kbps
- 384 kbps
- 768 kbps
- Auto (dynamically determined; supports any call speed in the range of 128 kbps to 768 kbps)

Note Each of the above speeds represents a multiple of 64 kbps. For 56-kbps dialing, there is a check-box on the service prefix configuration page to restrict each channel to 56 kbps. Therefore, a 128-kbps service with restricted mode enabled would result in a 112-kbps service; a 384 kbps service with restricted mode enabled would result in a 336-kbps service; and so on.

Calls from an IP endpoint toward the PSTN must include the service prefix at the beginning of the called number in order for the gateway to decide which service to use for the call. Optionally, you can configure the default prefix to be used for calls that do not include a service prefix at the beginning of the number. This method can become quite complex because users will have to remember which prefix to dial for the speed of the call they wish to make, and you would have to configure multiple route patterns in Unified CM (one for each speed). Fortunately, the Auto speed enables you to minimize this effort. If the majority of your calls are made using 64 kbps per channel (for example, 128 kbps, 384 kbps, 512 kbps, 768 kbps, and so on), you could use the Auto service in that case. You would then need to create only one other service for the rare case in which someone makes a call using 56 kbps per channel (for example, 112 kbps, 336 kbps, and so on).

Cisco recommends that you always use a # character in your service prefixes because the gateway recognizes the # as an end-of-dialing character. By placing this character in the service prefix, you block people from attempting to use the gateway for toll fraud by dialing the main number of the gateway, reaching the IVR, and then dialing out to an off-net number. The # can either be at the beginning (recommended) or the end of the service prefix. For example, if your access code to reach the PSTN is 8 for video calls, Cisco recommends that you configure the service prefix as #8 or 8#. Or, if you have two service prefixes as described above, you might use #80 for the Auto 64-kbps service and #81 for the Auto 56-kbps service.

The ramification of using a service prefix is that Unified CM must prepend the service prefix to the called number when sending calls to the Cisco Unified Videoconferencing gateway. Because forcing users to dial the # would not be very user-friendly, Cisco recommends that you configure Unified CM to prepend the # to the dialed number. For example, if the access code to dial a video call to the PSTN is 8, you could configure a route pattern as 8.@ in Unified CM, and in the route pattern configuration you would configure the called number translation rule to prepend #8 whenever that route pattern is dialed. Or, if you have two service prefixes as described above, you might use 80.@ for the Auto 64-kbps service (prefixing # to the called number) and 81.@ for the Auto 56-kbps service (prefixing # to the called number).

Automated Alternate Routing (AAR)

When the IP network does not have enough bandwidth available to process a call, Unified CM uses its call admission control mechanism to determine what to do with the call. As described in the chapter on IP Video Telephony, page 12-1, Unified CM performs one of the following actions with the call, depending on how you have configured it:

- Fail the call, playing busy tone to the caller and displaying a Bandwidth Unavailable message on the caller's screen
- Retry the video call as an audio-only call
- Use automated alternate routing (AAR) to re-route the call over an alternative path, such as a PSTN gateway

The first two options are covered in the chapter on IP Video Telephony, page 12-1, and this section covers the AAR option.

To provide AAR for voice or video calls, you must configure the calling and called devices as members of an AAR group and configure an External Phone Number Mask for the called device. The External Phone Number Mask designates the fully qualified E.164 address for the user's extension, and the AAR group indicates what digits should be prepended to the External Phone Number Mask of the called device in order for the call to route successfully over the PSTN.

For example, assume that user A is in the San Jose AAR group and user B is in the San Francisco AAR group. User B's extension is 51212, and the External Phone Number Mask is 6505551212. The AAR groups are configured to prepend 91 for calls between the San Jose and San Francisco AAR groups. Thus, if user A dials 51212 and there is not enough bandwidth available to process the call over the IP WAN between those two sites, Unified CM will take user B's External Phone Number Mask of 6505551212, prepend 91 to it, and generate a new call to 916505551212 using the AAR calling search space for user A.

This same logic applies to video calls as well, with one additional step in the process. For video-capable devices, there is field called Retry Video Call as Audio. As described in the chapter on IP Video Telephony, page 12-1, if this option is enabled (checked), Unified CM does not perform AAR but retries the same call (that is, the call to 51212) as a voice-only call instead. If this option is disabled (unchecked), Unified CM performs AAR. By default, all video-capable devices in Unified CM have the Retry Video Call as Audio option enabled (checked). Therefore, to provide AAR for video calls, you must disable (uncheck) the Retry Video Call as Audio option. Additionally, if a call admission control policy based on Resource Reservation Protocol (RSVP) is being used between locations, the RSVP policy must be set to Mandatory for both the audio and video streams.

Furthermore, Unified CM looks at only the called device to determine whether the Retry Video Call as Audio option is enabled or disabled. So in the scenario above, user B's phone would have to have the Retry Video Call as Audio option disabled in order for the AAR process to take place.

Finally, devices can belong to only one AAR group. Because the AAR groups determine which digits to prepend, AAR groups also influence which gateway will be used for the rerouted call. Depending on your choice of configuration for outbound call routing to the PSTN, as discussed in the previous section, video calls that are rerouted by AAR might go out a voice gateway instead of a video gateway. Therefore, carefully construct the AAR groups and the AAR calling search spaces to ensure that the correct digits are prepended and that the correct calling search space is used for AAR calls.

While these considerations can make AAR quite complex to configure in a large enterprise environment, AAR is easier to implement when the endpoints are strictly of one type or the other (such as IP Phones for audio-only calls and systems such as the Tandberg T-1000 dedicated for video calls). When endpoints are capable of both audio and video calls (such as Cisco Unified Video Advantage or a Cisco IP Video Phone 7985G), the configuration of AAR can quickly become unwieldy. Therefore, Cisco recommends that large enterprise customers who have a mixture of voice and video endpoints give careful thought to the importance of AAR for each user, and use AAR only for select video devices such as dedicated videoconference rooms or executive video systems. Table 13-4 lists scenarios when it is appropriate to use AAR with various device types.

Table 13-4 When to Use AAR with a Particular Device Type

Device Type	Device is used to call:	Enable AAR?	Comments
IP Phone	Other IP Phones and video-capable devices	Yes	Even when calling a video-capable device, the source device is capable of audio-only, thus AAR can be configured to route calls out a voice gateway.
IP Phone with Cisco Unified Video Advantage, or Cisco IP Video Phone 7985G	Other video-capable devices only	Yes	Because the device is used strictly for video calls, you can configure the AAR groups accordingly.
	IP Phones and other video-capable devices	No	It will be difficult to configure the AAR groups to route audio-only calls differently than video calls.
Sony or Tandberg SCCP endpoint	Other video-capable devices only	Yes	Because the device is used strictly for video calls, you can configure the AAR groups accordingly.
	IP Phones and other video-capable devices	No	It will be difficult to configure the AAR groups to route audio-only calls differently than video calls.
H.323 or SIP client	Other video-capable devices only	Yes	Because the device is used strictly for video calls, you can configure the AAR groups accordingly.
	IP Phones and other video-capable devices	No	It will be difficult to configure the AAR groups to route audio-only calls differently than video calls

Least-Cost Routing

Least-cost routing (LCR) and tail-end hop-off (TEHO) are very popular in VoIP networks and can be used successfully for video calls as well. In general, both terms refer to a way of configuring the call routing rules so that calls to a long-distance number are routed over the IP network to the gateway closest

to the destination, in an effort to reduce toll charges. (For Cisco Unified CM Release 4.1, LCR basically means the same thing as TEHO.) Unified CM supports this feature through its rich set of digit analysis and digit manipulation capabilities, including:

- Partitions and calling search spaces
- Translation patterns
- Route patterns and route filters
- Route lists and route groups

Configuring LCR for video calls is somewhat more complicated than for voice calls, for the following reasons:

- Video calls require their own dedicated gateways, as discussed previously in this chapter
- Video calls require much more bandwidth than voice calls

With respect to dedicated gateways, the logic behind why you might or might not decide to use LCR for video calls is very similar to that explained in the section on Automated Alternate Routing (AAR), page 13-32. Due to the need to have different types of gateways for voice and video, it can become quite complex to configure all the necessary partitions, calling search spaces, translation patterns, route patterns, route filters, route lists, and route groups needed for LCR to route voice calls out one gateway and video calls out another.

With respect to bandwidth requirements, the decision to use LCR depends on whether or not you have enough available bandwidth on your IP network to support LCR for video calls to/from a given location. If the current bandwidth is not sufficient, then you have to determine whether the benefits of video calls are worth the cost of either upgrading your IP network to make room for video calls or deploying local gateways and routing calls over the PSTN. For example, suppose you have a central site with a branch office connected to it via a 1.544-Mbps T1 Frame Relay circuit. The branch office has twenty video-enabled users in it. A 1.544-Mbps T1 circuit can handle at most about four 384-kbps video calls. Would it really make sense in this case to route video calls up to the central site in order to save on toll charges? Depending on the number of calls you want to support, you might have to upgrade your 1.544-Mbps T1 circuit to something faster. Is video an important enough application to justify the additional monthly charges for this upgrade? If not, it might make more sense to deploy a Cisco Unified Videoconferencing gateway at the branch office and not bother with LCR. However, placing local Cisco Unified Videoconferencing gateways at each branch office is not inexpensive either, so ultimately you must decide how important video-to-PSTN calls are to your business. If video is not critical, perhaps it is not worth upgrading the bandwidth or buying video gateways but, instead, using the Retry Video Call as Audio feature to reroute video calls as voice-only calls if they exceed the available bandwidth. Once a call is downgraded to voice-only, local gateway resources and bandwidth to perform LCR become more affordable and easier to configure.

ISDN B-Channel Binding, Rollover, and Busy Out

With Cisco IOS Release 12.4.20T or later releases, Cisco IOS H.320 gateways support the ISO-13871 bonding technique, which supports video calls at speeds up to 1 Mbps for video calls. With this functionality the Cisco IOS router can be used as an integrated gateways for both voice and video calls.

H.320 video uses multiple ISDN channels bound together to achieve the speeds needed to pass full-motion video. One of the problems with this bonding mechanism is that, when an inbound ISDN video call is received, the gateway does not know how many channels will be requested for that call until after it accepts the call and the source device indicates how many additional channels are required. If there are not enough B-Channels to satisfy the request, the call is disconnected. Therefore, careful traffic engineering is required to minimize the possibility that this situation will occur. Essentially, you want to ensure that there are always enough B-Channels available to handle the next call that might come in.

This B-Channel issue occurs in two cases:

- Inbound calls from the PSTN to the IP network
- Outbound calls from the IP network to the PSTN

Inbound Calls

For inbound calls, consider the following scenario:

> A company has a Cisco Unified Videoconferencing 3527 Gateway with an ISDN PRI circuit connecting it to a central office (CO) switch. The ISDN PRI circuit in this case offers 23 B-Channels. A video call is received from the PSTN at 384 kbps. This call takes six B-Channels, leaving 17 available. A second and third 384-kbps call are received on the line while the first one is still active. These each take an additional six channels, leaving five channels available. When the fourth 384-kbps call is received, the gateway will answer the call but, recognizing that it does not have enough B-Channels available (it only has five left but the call requires six), it will disconnect (by sending a Q.931 RELEASE COMPLETE with "16: Normal Call Clearing" as the reason). The caller attempting to make the fourth call will not know why the call failed and might redial the number repeatedly, trying to make the call work.

On Cisco Unified Videoconferencing gateways, you can minimize your chances of running into this issue by configuring the gateway to send a request to the CO to busy-out the remaining B-Channels (in this example, five channels) whenever the gateway reaches a certain threshold of utilization (configured as a percentage of total bandwidth).

In addition, you can have the CO provision multiple ISDN circuits in a trunk group. When the first circuit reaches the busy-out threshold, calls will roll over to the next PRI in the group. The Cisco Unified Videoconferencing 3500 Series Gateway offers two ISDN PRI connections and supports bonding channels across both ports. For example, port 1 might have only five channels available while port 2 is sitting idle and, therefore, has 23 channels available. By taking the five channels from port 1 and one channel from port 2 and bonding them together, the fourth 384-kbps call can succeed. This leaves 22 channels available on controller 2, and at some point additional inbound calls would reach the busy-out threshold again. At that point the remaining channels on port 2 will be busied out, and all further inbound calls will be rejected with cause code "Network Congestion." Cisco Unified Videoconferencing gateways cannot bound channels across different gateways or across different Cisco 3500 Series gateway models in the same Cisco 3545 chassis, so two ports is the maximum that you can bond together. The CO switch can still roll calls over to a third or forth PRI in the trunk group (most COs support trunk groups of up to 6 circuits), but you cannot bond channels between PRI number one and PRI number three, for example, as you can between PRI number one and PRI number two.

The busy-out logic described above depends on the assumption that all calls take place at the same speed. Suppose, for example, that two 384-kbps calls are active on a port and a 128-kbps call came in. This call would take only two channels, using a total of 14 channels for the three calls (6+6+2 = 14) and leaving nine channels available on the circuit. However, if the busy-out threshold is set at 18 channels (assuming that all calls would take place at 384-kbps), only four channels are still available under this busy-out threshold. If another 384 kbps call comes in at this point, the call will fail because the remaining four channels are not enough to support the call. Also, because the busy-out threshold of 18 channels has not been reached yet (only 14 channels are used), the circuit is not busied out and calls will not roll over to the next circuit. This condition will persist until one of the existing calls is disconnected. To avoid such situations, it is important to try to standardize on a single call speed for all calls.

Outbound Calls

Outbound calls encounter the same potential situations as inbound calls, but the way in which the busy-out occurs is different. The Cisco Unified Videoconferencing 3500 Series Gateways support messages called Resource Availability Indicator and Resource Availability Confirm (RAI/RAC). The RAI/RAC messages are defined under the H.225 RAS specification and are used by the gateways to tell the gatekeeper that they are full and to no longer route any more calls to them. When the gateway reaches the busy-out threshold, it sends an RAI message with a status of True to the gatekeeper. True means "Do not send me any more calls;" False means "I am available." The gateway sends an RAI=False as soon as it is no longer at its busy-out threshold. The busy-out threshold for outbound calls is separate from the busy-out threshold for inbound calls, and you can configure them differently so that inbound calls will roll over to the next available circuit but outbound calls will still be accepted, or vice versa. For example, you could configure the RAI threshold to 12 channels but the ISDN busy-out threshold to 18 channels. When two 384 kbps are active, outbound calls will roll over to the next available gateway, but a third 384-kbps inbound call could still be received. An equally efficient method of achieving outbound call busy-out failover is to use Unified CM's route group and route list construct, as described in the following section, instead of the RAI/RAC method.

Configuring the Gateways in Unified CM

You can configure a Unified Videoconferencing gateway in either of the following ways in Unified CM:

- Configure it as an H.323 gateway, and Unified CM will route calls directly to the gateway.
- Configure an H.225 gatekeeper-controlled trunk to the gatekeeper, and route calls to the gateway through the gatekeeper.

If you have only one gateway, it is probably easier to configure it directly in Unified CM instead of going through a trunk to get to it. If you have multiple gateways for load balancing and redundancy, you can either configure them all in Unified CM and place them into a route group(s) and route list, or configure an H.225 trunk to the gatekeeper and rely on RAI/RAC between the gateways and the gatekeeper to tell Unified CM which gateway it should send a given call to.

For inbound calls from the PSTN to Unified CM, the Cisco Unified Videoconferencing gateways can either register with a gatekeeper or be configured with the IP addresses of up to three Unified CM servers to which they should send all inbound call requests. This method is known as peer-to-peer mode. Either way, the goal is have all inbound calls received by the gateways sent to Unified CM so that Unified CM can decide how to route the calls. See Gatekeepers, page 12-25, for more details on how to configure the gatekeeper to route calls from the gateways to Unified CM.

Call Signaling Port Numbers

By default, the Cisco Unified Videoconferencing Gateways listen on TCP port 2720 instead of the well-known port 1720. However, also by default, Unified CM sends H.323 calls to port 1720. You can change the port that the gateway listens on or you can change the port that Unified CM sends to in the H.323 gateway device configuration in Unified CM. Either way, both sides have to match in order for outbound calls to the gateway to succeed.

In the inbound direction, when configured to operate in peer-to-peer mode, the Cisco Unified Videoconferencing Gateways will send the call to Unified CM on port 1720. When configured to register with a gatekeeper, Unified CM uses a randomly generated port number for all gatekeeper-controlled trunks. This method enables Unified CM to have multiple trunks to the same gatekeeper. This port number is included in the Registration Request (RRQ) from Unified CM to the gatekeeper, so the inbound H.225 setup message from the gateway to Unified CM will be sent to this

port number. However, if the gateway is configured directly in Unified CM as an H.323 gateway device, Unified CM will ignore the fact that the call came in on the TCP port of the H.225 trunk and will instead match the source IP address to the H.323 gateway device configured in its database. If it does not find a matching device, Unified CM will treat the call as if it came in on the trunk.

In the outbound direction, if Unified CM uses a gatekeeper-controlled H.225 trunk to reach the gateway, the gatekeeper will tell Unified CM which TCP port to use to reach the gateway. If the gateway is configured in Unified CM as an H.323 gateway device (that is, peer-to-peer mode), then Unified CM must be configured to send calls either to port 2720 (default) or to 1720 (if the listening port on the gateway has been modified).

Call Signaling Timers

Due to the delay inherent in H.320 bonding, video calls can take longer to complete than voice calls. Several timers in Unified CM are tuned, by default, to make voice calls process as fast as possible, and they can cause video calls to fail. Therefore, you must modify the following timers from their default values in order to support H.320 gateway calls:

- H.245TCSTimeout
- Media Exchange Interface Capability Timer
- Media Exchange Timer

Cisco recommends that you increase each of these timers to 25 by modifying them under the Service Parameters in Unified CM Administration. Note that these are cluster-wide service parameters, so they will affect calls to all types of H.323 devices, including voice calls to existing H.323 Cisco Voice Gateways.

Bearer Capabilities of Voice Gateways

H.323 calls use the H.225/Q.931 Bearer Capabilities Information Element (bearer-caps) to indicate what type of call is being made. A voice-only call has its bearer-caps set to "speech" or "3.1 KHz Audio" while a video call has its bearer-caps set to "Unrestricted Digital Information." Some devices do not support Unrestricted Digital Information bearer-caps. Calls to these devices might fail if Unified CM attempts the call as a H.323 video call.

Unified CM decides which bearer-caps to set, based on the following factors:

- Whether the calling and/or called devices are video-capable
- Whether the region in Unified CM is configured to allow video for calls between those devices

Unified CM supports retrying the video call as audio, and this feature can be enabled through configuration. When Unified CM makes a video call with bearer-caps set to "Unrestricted Digital" and the call fails, Unified CM then retries the same call as an audio call with the bearer-caps set to "speech."

When using H.323, Cisco IOS gateways can service calls as voice or video, based on the bearer capabilities it receives in the call setup. When using SIP, the gateway translates the ISDN capabilities into SDP for call negotiations.

If the Cisco voice gateway uses MGCP to communicate with Unified CM, the problem will not occur because Unified CM does not support video on its MGCP protocol stack and because, in MGCP mode, Unified CM has complete control over the D-Channel signaling to the PSTN.